USA **89**:6721-6725 (1992); Takeuchi et al., Genes Dev. **9**:1211-1222 (1995); and Takahashi et al., Science **264**:1724-1733 (1994)). The gene trap approach is limited to genes expressed in ES cells, although variations of the method have been developed for targeting specific subclasses of genes expressed in early embryonic stages (Wurst et al., Genetics **139**:889-899 (1995); Skarnes et al., Proc. Natl. Acad. Sci. USA **92**:6592-6596 (1995); and Forrester et al., Proc. Natl. Acad. Sci. USA **93**:1677-1682 (1996)). And the chemical/radiation induced mutagenesis technique is generally limited to genes that can result in dominant phenotypes when mutated. None of these approaches, as currently exploited, may be readily streamlined or automated, nor can they be readily adapted to carry out saturated mutagenesis of the mouse genome.

Summary of the Invention

In general, the invention features a method for mutagenizing a mammalian gene, the method involving introducing into a mammalian cell (for example, a stem cell, such as an embryonic stem cell) a retroviral vector, the vector including a splice acceptor sequence, a transcription termination sequence, and retroviral packaging and integration sequences, the introducing step being carried out under conditions which allow the vector to integrate into the genome of the cell.

In preferred embodiments, the retroviral vector includes packaging and integration sequences derived from a Moloney murine leukemia virus sequence; the retroviral vector further includes a reporter gene whose expression is under the control of a mammalian cell promoter, the promoter being operably linked to the reporter gene upon integration of the vector into the genome of the mammalian cell; the reporter gene encodes a regulatory protein, the regulatory protein being capable of modulating the expression of a detectable gene; the regulatory protein is a tetracycline repressor fused to an activator protein (for example, VP16); the retroviral vector further includes a DNA sequence encoding a constitutively expressed marker gene, the marker gene being detectable in a mammalian

cell; the marker gene is a green fluorescent protein (for example, a green fluorescent having increased cellular fluorescence relative to a wild type green fluorescent protein); the green fluorescent protein is fused to a mammalian selectable marker; the mammalian selectable marker encodes neomycin resistance; the retroviral vector further includes a recognition sequence derived from a yeast VDE DNA endonuclease; the retroviral vector further includes a sequence which is recognized by a recombinase enzyme (for example, a loxP sequence); the mammal is a mouse; and the cell is an embryonic stem cell.

In a related embodiment, the invention features a retroviral vector which includes a splice acceptor sequence, a transcription termination sequence, and retroviral packaging and integration sequences. In preferred embodiments, the retroviral vector includes packaging and integration sequences derived from a Moloney murine leukemia virus sequence; the retroviral vector further includes a reporter gene whose expression is under the control of a mammalian cell promoter, the promoter being operably linked to the reporter gene upon integration of the vector into the genome of the mammalian cell; the reporter gene encodes a regulatory protein, the regulatory protein being capable of modulating the expression of a detectable gene; the regulatory protein is a tetracycline repressor fused to an activator protein (for example, VP16); the detectable gene includes an operably linked tetracycline operator; the retroviral vector further includes a DNA sequence encoding a constitutively expressed marker gene, the marker gene being detectable in a mammalian cell; the marker gene is a green fluorescent protein (for example, a green fluorescent protein having increased cellular fluorescence relative to a wild type green fluorescent protein); the green fluorescent protein is fused to a mammalian selectable marker; the mammalian selectable marker encodes neomycin resistance; the retroviral vector further includes a recognition sequence derived from a yeast VDE DNA endonuclease; and the retroviral vector further includes a sequence which is recognized by a recombinase enzyme (for example, a loxP sequence).

In other related embodiments, the invention includes a cell containing a retroviral

vector of the invention; a transgenic non-human mammal (for example, a mouse) which includes a retroviral vector of the invention; a library (that is, having at least 100 members) of mutagenized mammalian genes produced by the methods of the invention; and cells (for example, stem cells) which include a library of mutagenized mammalian genes produced by the methods of the invention.

In a related method, the invention features a method for identifying a cell (for example, a stem cell) which includes a retroviral vector, the method involving: (a) introducing into a mammalian cell population a retroviral vector, the vector including a splice acceptor sequence, a transcription termination sequence, retroviral packaging and integration sequences, and a constitutively expressed detectable marker gene, the introducing step being carried out under conditions which allow the vector to integrate into the genomes of the cells; and (b) identifying the cell which includes the retroviral vector by detecting expression of the marker gene.

In preferred embodiments, the marker gene is a green fluorescent protein; and the green fluorescent protein has increased cellular fluorescence relative to the wild-type green fluorescent protein.

In a second related method, the invention features a method for identifying a mutagenized mammalian gene, the method involving: (a) introducing into a mammalian cell (for example, a stem cell) population a retroviral vector, the vector including a splice acceptor sequence, a transcription termination sequence, and retroviral packaging and integration sequences, the introducing step being carried out under conditions which allow the vector to integrate into the genomes of the cells; (b) isolating the genomic DNA from the population of cells; (c) amplifying the genomic DNA using amplification primers based at least in part on the retroviral sequence; and (d) identifying the mutagenized mammalian gene by sequence homology with a wild-type nucleic acid sequence. In a preferred embodiment, the sequence homology is identified using a hybridization technique.